

Restriction Enzyme Databases - Thanks to Dr. Richard Roberts, one of the members of the BIONET National Advisory Committee, we have established one of the most up-to-date lists of restriction enzymes available. Dr. Roberts maintains a restriction enzyme registry and distributes his updated lists on a monthly basis in an electronic message to BIONET. BIONET brings these new lists into use within about 24 hours. These lists are used within the core programs SEQ and PEP and are referenced 592 times per month. Recently Dr. Robert's group has included with each update a description of the new enzymes and new data about existing enzymes. This material is available on-line.

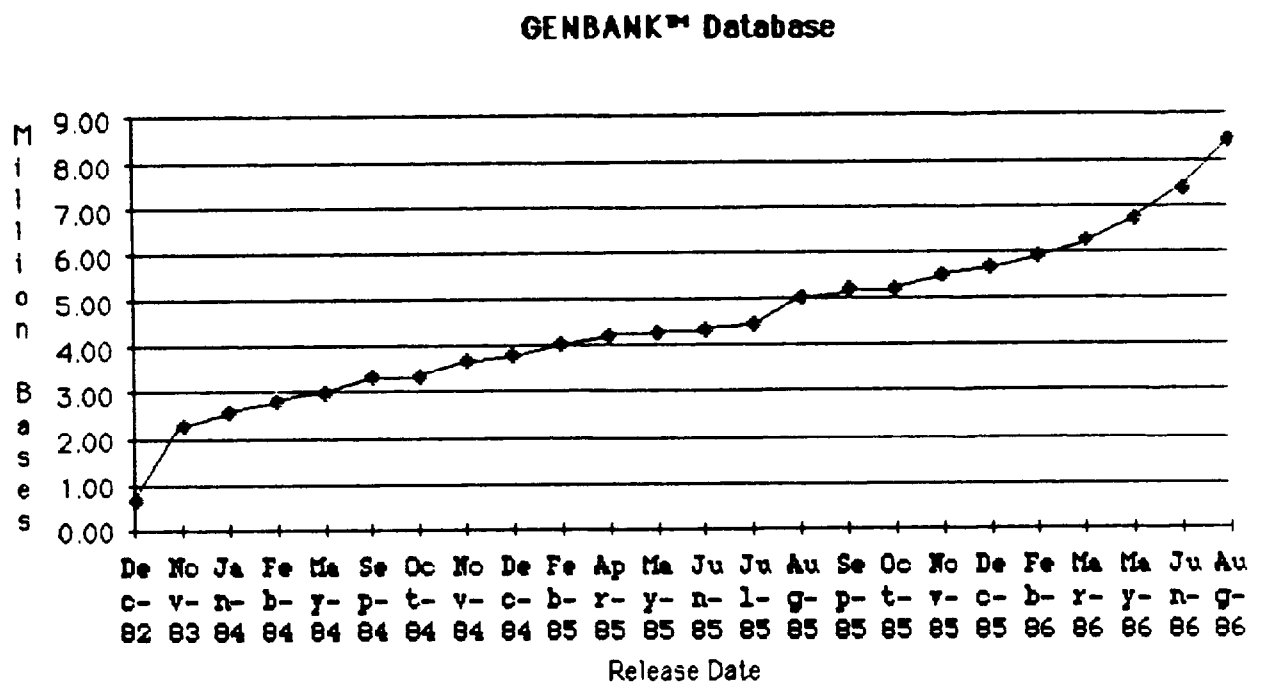
Vectorbank - Vectorbank is a collection of restriction maps of important cloning vectors, viruses and phages that is maintained by the IntelliGenetics staff. This database is used by the CLONER program for manipulating restriction maps and simulating DNA cloning experiments. Initially many of the restriction maps in Vectorbank were taken directly from sequence information in the sequence databases. The current database contains many restriction maps contributed directly by BIONET users or by scientists that have developed the vectors for various uses (sequencing vectors, expression vectors etc.).

We also contacted Elsevier who have recently published the book *Cloning Vectors*, containing restriction maps of over a thousand vectors and phages, but unfortunately the information in that book is not available in a computer readable form. Nevertheless, the book provides an important resource and as individual scientists enter the map data into the computer, either through the Cloner or Strategene programs, we can make them available to the entire BIONET community.

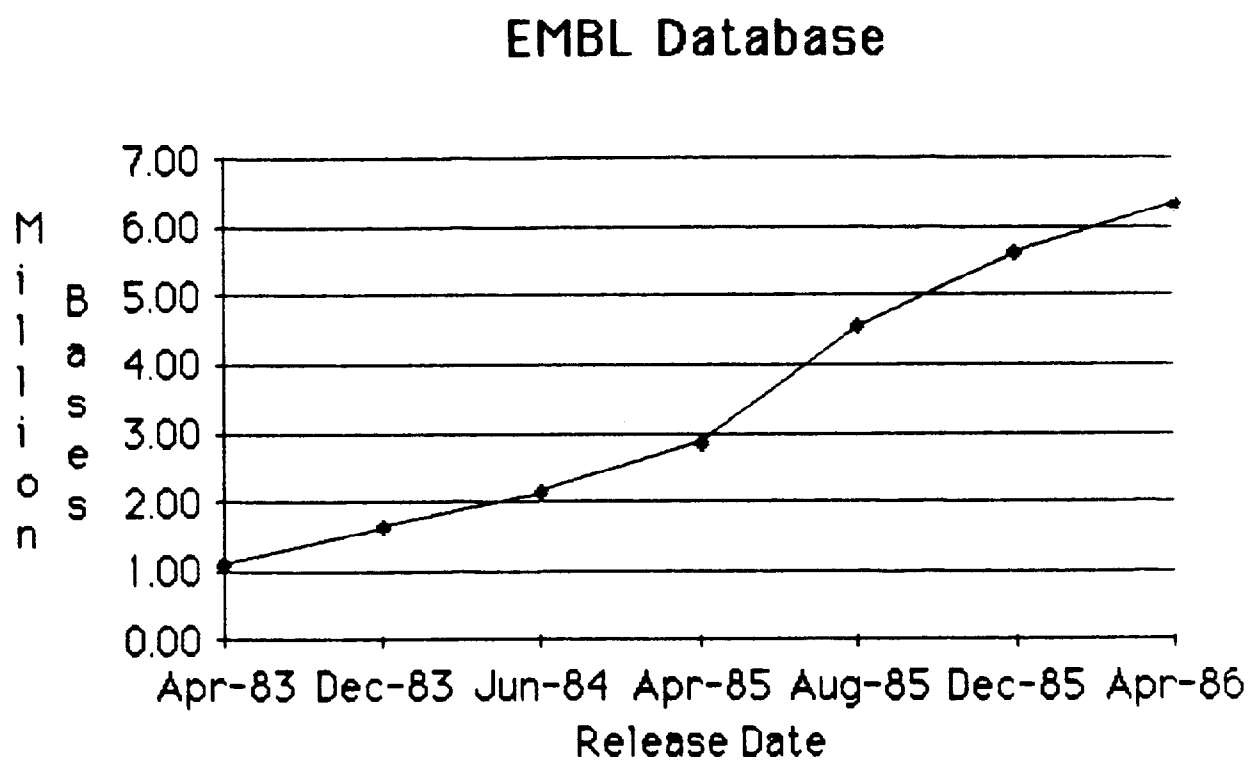
The vectorbank is referenced about 70 times per month and the CLONER program itself is used 131 times per month.

Consensus Sequence Oligonucleotide Databank - Recently Drs. Trifonov and Brendel have published a compendium called *Gnomic* of over 800 consensus sequences from the literature. This book contains a list of short oligonucleotide sequences which are either tandem repeats, promoter sequences, operators, protein binding sites, restriction enzyme cleavage sites, or other short sequences of functional importance. Dr. Trifonov has already made this list available to us and has requested his publisher to make the entire book available in a computer readable form. We intend to edit this list of sequences into a set of KEYs that can be used for database searching via the BIONET QUEST program. We will call this database a "KEYBANK". In addition to the sequences themselves we intend to include the literature references that reported the sequence and a cross reference to the GNOMIC book. The QUEST program is already used for this purpose but the collection of KEY's that we have is quite small and is collected in a non-systematic fashion. QUEST is currently among the most heavily used programs with 864 references per month and with an extensive databank of keys we expect that usage to increase.

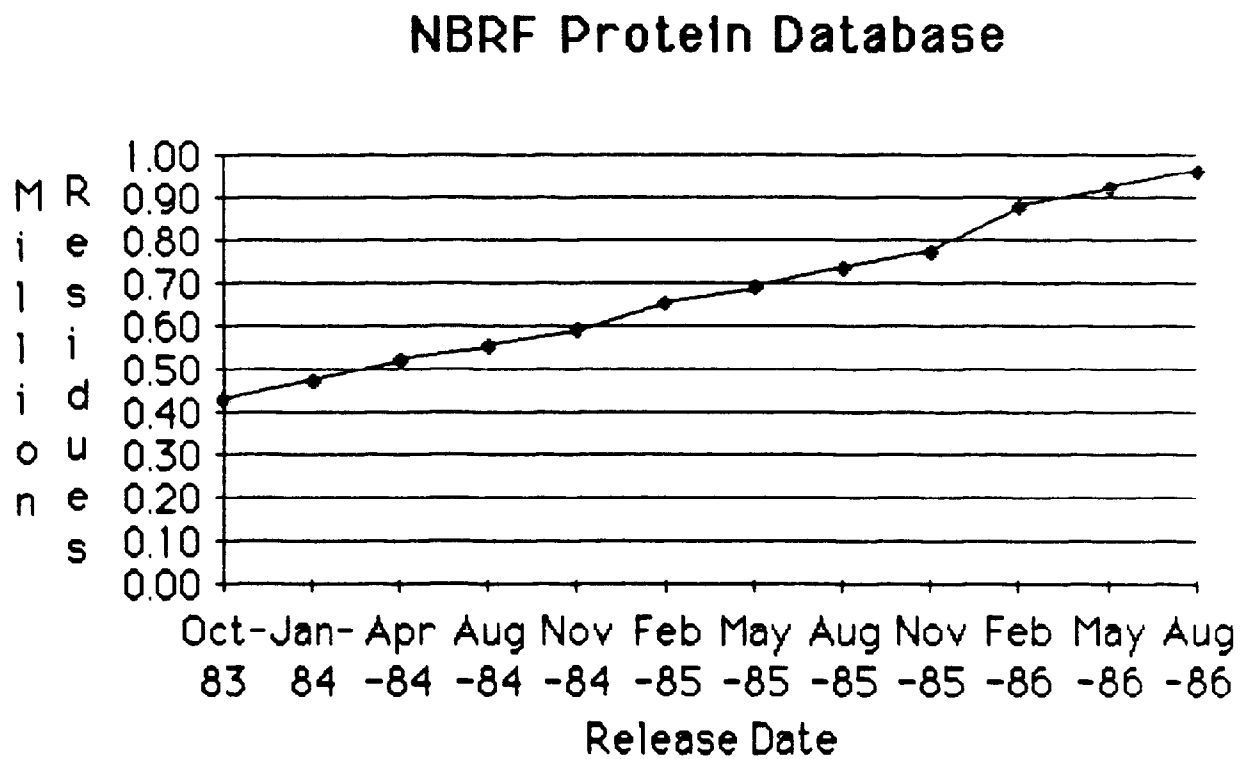
Genetic Map Of *Drosophila melanogaster* - Last year we made a computer readable version of Lindsley

**Figure III-8: Size and Release Dates of the NIH GenBank Database**

The size and release dates of the NIH GenBank database on the BIONET computer. This figure includes a few releases prior to the BIONET grant.

**Figure III-9: Size and Release Dates of the EMBL DNA Sequence Database**

The size and release dates of the EMBL DNA sequence database

**Figure III-10: Size and Release Dates of the PIR Protein Database**

The size and release dates of the Protein Identification Resource protein database maintained by the National Biomedical Research Foundation.

and Grell's *Genetic Variations of Drosophila melanogaster* available on line. We described how to search this text for genetic markers and rearrangements, and, most importantly, we described how to find all genetics markers within a specified region. Despite the fact that only 15 BIONET scientists work with *Drosophila*, this database was referenced 79 times for a rate of about 10 references per month.

Dr. Lindsley is currently updating this 18 year old reference work which has more than trebled in size. During this year we have mounted two updates to this work and now have the entire list of mutants as of 1986 available for searching. In addition we are working with Dr. John Merriam at UCLA who is preparing a molecular correspondence between cloned gene segments and the genetic map. Dr. Merriam's work is a database of all mapped cloned sequences and provides an important reference for starting cloning experiments in this organism.

*Drosophila* is a very good organism for presenting the entire genetic map on BIONET because the data is completely available in a computer readable form. As more and more genetic maps become available we will consider the best way to provide them to the BIONET community. Dr. Frank Ruddle has also made the genetic map of humans available as a Spires database at Yale. We will attempt further collaborations where this seems useful and feasible.

Brookhaven Protein Structure Database - Although we have no programs for displaying three dimensional protein structure on the BIONET computer, there are a number of BIONET users that have access to both mainframe and more recently, microcomputer programs for displaying and editing this type of information. Several of our users have requested that we make this type of information available. We have obtained the database from the Brookhaven Laboratory and mounted its files on BIONET. While the entire database is too large for BIONET users to download into their microcomputer, it should be a valuable resource for those wishing to obtain a single protein structure or those wanting the most recent updates to their own library. We are considering other ways of making this three dimensional data of value to the BIONET community.

### III.B. Highlights

Total usage on the system has increased by 25% over the last year. About 170 inactive accounts were replaced by 130 active new users.

We have taken several steps to improve the electronic communications available through BIONET, including substantial work on revamping the bulletin board system and installation of mail forwarding among facilities via Telenet and ARPANET.

We have begun the collection and dissemination of several programs designed to help solve the very

important problem of alignment of multiple biological sequences. This important Core Research activity will place BIONET in a leadership role in making such programs routinely available.

A new system of on-line help menus was installed to assist scientists in using the resource.

The Brookhaven Protein Databank is now available on BIONET along with a new facility designed to expedite dissemination of sequence data.

### **III.C. Administrative Changes**

There have been many administrative changes within BIONET during the past year. These have come about for reasons ranging from a change in ownership of IntelliGenetics itself, to personnel shifts, additions, and resignations. We are fortunate that none of these changes has had a significant impact on the Resource itself, in particular, its appearance and availability to the community. Here is a brief summary of changes:

- In early May, 1986, IntelliGenetics became a Joint Venture between Amoco and IntelliCorp, 60% owned by Amoco and 40% by IntelliCorp. As part of this transition, the management of the Cooperative Agreement was turned over completely to IntelliGenetics, and Dr. Dennis H. Smith replaced Dr. Ralph Kromer as Principal Investigator.
- In late April, 1986, Dr. Marcia Allen took a leave of absence from her position as a BIONET Scientific Consultant; this position was filled in early August, 1986 by Dr. David Kristofferson. Dr. Kristofferson has recently been appointed Acting Resource Manager.
- Mr. David Horner joined the Computer Facilities Group as a Senior Operator in early June, 1986, replacing Ms. Mary Yardley, who transferred to another position in IntelliGenetics. Mr. Horner spends 50% of his time on BIONET-related tasks concerning the operation of the DEC-2060.
- On July 1, 1986, Ms. Nancy Bigham, a BIONET Scientific Consultant, transferred to IntelliGenetics Customer Support, concurrent with the transfer of Ms. Theresa Friedemann from Customer Support to BIONET Scientific Consultant.
- Mr. Rob Liebschutz joined the Computer Facilities Group as a systems programmer in early November, and is spending about 50% of his time with BIONET working on BIONET Satellite and ARPANET communications.
- Dr. Sunil Maulik joined IntelliGenetics in early November, 1986, and is working 100% time as a BIONET Scientific Consultant.
- Dr. Dennis Smith resigned from IntelliGenetics on December 1, 1986 to pursue other opportunities. His position as Principal Investigator has been taken by Dr. Michael Kelly, President of IntelliGenetics.

The resignation of Dr. Smith is the only administrative change that might temporarily affect the operation of BIONET. However, as part of his transition, arrangements have been made for commitment

of additional time of the BIONET Co-Investigators Drs. Brutlag, Friedland and Kedes. This will ensure a smooth transition while a search for additional senior staff for BIONET takes place.

### **III.D. Resource Advisory Committee and Allocation of Resources**

The membership of BIONET's National Advisory Committee remains unchanged from last year. It consists of:

- Professor Joshua Lederberg, MD, PhD. (Chair), President, The Rockefeller University.
- Dr. Saul Amarel, PhD., Director, Information Systems Technologies Office, Defense Advanced Research Projects Agency, Department of Defense.
- Professor Alan Maxam, PhD., Dana Farber Cancer Institute, Harvard Medical School, Harvard University.
- Dr. Richard J. Roberts, PhD., Senior Staff Investigator, Molecular Biology, Cold Spring Harbor Laboratory
- Thomas Rindfleisch, MS, Director, Knowledge Systems Laboratory, Department of Computer Science, Stanford University.
- Professor Charles Yanofsky, PhD., Department of Biological Sciences, Stanford University.
- Professor Fotis Kafatos, PhD., Department of Cellular and Developmental Biology, Harvard University.

Meetings of the Committee are normally held once a year. The last regularly scheduled meeting occurred at the end of the previous grant year, on February 24, 1986 in Mountain View. At this meeting we reviewed the progress of the Resource and discussed directions for Core and Collaborative Research for the current grant year covered by this Report. In particular, we discussed our goals for pursuing three Core Research projects, multiple sequence alignment, network communications associated with the BIONET Satellite Program, and special hardware for rapid text searching. All three projects were given the strong endorsement of the Committee; progress in the three areas was summarized previously under *Core Research*, section III.A.3, above.

Another strong recommendation of the Committee was that BIONET should be truly international in scope, and that efforts to extend communications and collaborations to other resources in foreign countries should be encouraged. We have been hampered in carrying out this recommendation because of delays in the installation of ARPANET, but this will be completed shortly. When installed, we will immediately implement electronic mail facilities and routing directories so that communication among international sites is made simple. Several such sites were mentioned previously under *Collaborative Research*, section III.A.3 above.

The Committee agrees with our methods for allocating the Resource. The DEC-2060 computer uses its windfall scheduler to allocate cpu time to the various categories of users and overhead, as described in our previous Report. The cpu time is distributed on a first-come, first-served basis. This method has been very successful, with considerably more than BIONET's 50% time allocation being delivered to BIONET scientists (see section III.A.5.b, above).

We continue to request that the community not have more than one person per PI group using BIONET at the same time during prime time. The community continues to do an excellent job in complying with this policy.

We continue to allocate additional disk space to PI groups involved in managing large sequencing projects or extensive databases of sequences. We do this on an *ad hoc* basis upon requests by investigators. We project that sufficient disk space will be available through the remainder of this grant year, and probably beyond, to meet the storage requirements of the community. Our archive and retrieval system is working smoothly to store seldom-used files on tape and retrieve them promptly (1-2 days) upon requested.

### **III.E. Dissemination of Information of Resource's Capabilities**

We discuss two areas related to dissemination of information about the Resource that we have pursued this grant year. The first is interactions with the scientific community through participation at conferences. The second is use of the electronic mail and bulletin board facilities of the Resource itself to keep the BIONET community aware of changes and improvements.

#### **III.E.1. Community Interactions and Awareness**

We have used two methods this year to inform the community about BIONET and to solicit applications for access to the Resource. The first method has been participation at major conferences where we have presented papers and/or have had booths at exhibitions. These efforts are summarized previously under Training, Subsection III.A.4. At these conferences, we have distributed the standard applications packets to scientists, after demonstrating to them the capabilities of the Resource. (See Appendix VIII for an example of the renewal newsletter).

Last year we placed one advertisement in *Nucleic Acids Research* and while we did receive some response, most of the queries came from Soviet bloc countries. We received a greater response from our presence at the above mentioned conferences and focussed our advertising efforts in that direction. See Appendix IX for a copy of the advertisement which appeared in the Miami Mid-Winter Symposia, 1987 brochure.

This year was the first year that we solicited the renewal of subscription fees. We used this opportunity



to bring our subscribers up to date on the changes to the resource. A sample of the newsletter is in Appendix X.

### **III.E.2. Electronic Communications**

The electronic communication facilities of BIONET provide another important way to disseminate information about the Resource. In addition, electronic mail and bulletin boards provide a mechanism for scientific and technical interchanges among members of the community.

#### **III.E.2.a. Bulletin Boards**

The electronic bulletin boards are an important component of the BIONET Resource. They provide BIONET users with a facility for the exchange of data, laboratory techniques and ideas. Our users represent a wealth of knowledge. Communication is the key to accessing and disseminating that knowledge.

Information on the current status of the Bulletin Boards has been previously discussed in Section III.A.2.e. An announcement used to attract new bulletin board leaders is given in Appendix XI.

### **III.F. Suggestions and Comments**

In the previous year's Annual Report, we made two strong suggestions for improvement. The first was a plea for more warning about the effects of Federal budgetary decisions on our grant award. We are pleased that we had ample warning of potential cuts this year, which dramatically improved our own planning processes and ability to react. We are not pleased that cuts had to take place, but we did appreciate the warning.

Our second suggestion was related to the lack of NIH initiatives in computer networking, making it difficult for computer resources to share information electronically. We have discussed this problem at length with the BRTP staff and with our counterparts at other resources. The net result is general agreement that: 1) there are already many efforts aimed at coordinating the growth of computer networks; 2) a new effort may be counterproductive; and 3) the NIH in general, and BRTP in particular, should do all it can to encourage its grantees to participate in existing networks. We feel the best way to achieve (3) is to encourage persons submitting proposals to anticipate the costs of networking, and to ensure that Study Sections and Councils understand the importance of maintaining these funds in final awards.

# I. BIONET RESOURCES REAPPLICATION

9 December 1986

Dear BIONET Principal Investigator:

The National Institute of Health requires, as part of our Annual Report, that we review the status of all BIONET subscribers each year. Thus, we need information from you on any changes, from your original application, with respect to institutional affiliation, address, funding status, sub-investigators, etc. Most importantly, we need a list of all publications in which BIONET played a role; as well as three copies of the reprint or preprint. We also ask that at this time you reaffirm your original agreement for access to BIONET.

We are enclosing a printout of the record we now have on your lab. Please indicate on the printout any changes in:

- Your title, affiliation, mailing address and/or phone number
- The list of your sub-investigators who have a subdirectory in your account.

On page 1 of the BIONET Resource Reapplication please type or print your full name and title. Read and affix the date, name of official and signatures to the BIONET agreement.

On page 2, please type or print your name and note change in status of funding and provide a list of current publications resulting, in part, from the use of the BIONET Resource (Remember to cite the BIONET Grant # 1 U41 RR-01685-03 in all such publications.)

ON page 3, type or print your name and provide a brief description on how BIONET was used in your research.

We would also like to give you this opportunity to comment on the BIONET resource - what role it is playing in your research and any suggestions/requests for improvement.

Because we must prepare our Annual Report in December, we need you to return this re-application to us no later than November 7, 1986. Thank you for your cooperation.

Sincerely,

Mary Lou Warner  
BIONET Administrator

## **BIONET RESOURCE**

### **Reapplication**

**Fiscal 1987**

**Principal Investigator (full name and title):**

### **BIONET Agreement**

As Principal Investigator of this grant to use the BIONET Resources, I agree to adhere to all conditions and restrictions for use of the BIONET Resource, as described in the document "The BIONET<sup>tm</sup> Resource, Description and Applications Form" and such further regulations as may be issued from time to time by the NIH or BIONET's National Advisory Committee.

The BIONET Resource will not be used for any commercial purpose which is not specifically identified to and approved by the NAC. Any pertinent change in sponsorship, continuity of grant support, or use made of BIONET will be reported promptly to the BIONET Resource Manager.

I have also furnished a copy of this re-application to the Grants Administrator of my institution, whose signature appears below.

I also assume full responsibility for all users listed on this applications form and will monitor their compliance to the conditions and restrictions for access to the BIONET Resource. I will inform the BIONET Consultant, (electronic mail address: BIONET), by electronic mail, immediately about any changes in this group of users, i.e., departure of existing user or addition of new staff qualified to use the resource. I will inform new users of the above mentioned conditions and restrictions.

**Date:**

**Name of official:** \_\_\_\_\_

**Signature of Principal Investigator**

**Signature of Grants Administrator**

**BIONET re-application page 2****Funding Status**

Please note any new funding, including Institution, Grant Number, title of grant, and duration of grant.

**Current Publications**

List current publications resulting, in part, from the use of the BIONET Resource (use standard bibliographic format). Remember to cite the BIONET Grant # 1 U41 RR-01685-03 in all such publications. A sample citation would be: *Computer resources used to carry out our studies were provided by the N.I.H. sponsored BIONET<sup>tm</sup> National Computer Resource for Molecular Biology.* Please submit 3 copies of preprints or reprints. If you would like to have any of your other publications listed on the BIONET Bibliographies Bulletin Board, please list those on a separate sheet.

PI Name: \_\_\_\_\_

**BIONET Re-application    Page 3****Use of BIONET**

**Briefly describe how BIONET has helped your research:**

**COMMENTS**

We invite your comments, suggestions and requests about the BIONET Resource. Which programs are the most useful to you - the least? Should the bulletin boards be broader in scope - more specific? Would you like more interaction with other users? What else would you like to see included in the BIONET Resource, for example, other computer programs. Would you like more information about the BIONET on-site package?

## II. ON-LINE HELP MENUS

### ●HELP ME

WELCOME !!!!

This menu is designed to provide answers for common questions.  
Please send any comments about it by electronic mail to BIONET.

#### FOR INFORMATION ON:

#### TYPE AFTER THE @

Support Phone Numbers for BIONET users	HELP CONSULTANT
Support Phone Numbers for Commercial Users	HELP SAR
Making Printed Copies of Information and Files	HELP HARDCOPY
How to Use the Editors	HELP ED
How to Find Other Users on the System	HELP WHOIS
How to Use the Electronic Mail	HELP E-MAIL
How to Use Bulletin Boards (and SAVE \$400!!)	HELP BULLETINS
Scientific Meetings by Computer !!	HELP MEETINGS
A Guide to the Main IntelliGenetics Programs	HELP MAIN-PROGRAMS
Contributed Programs and FREE PC Software	HELP SOFTWARE
File Copying, Deleting, etc.; CTRL Key Use	HELP TOPS20
Transferring Files Between Computers	HELP FILE-TRANSFER
Nucleic Acid, Protein, and other Databases	HELP DATABASES
How to Find Sequence Files	HELP FIND-FILES
Running Time-Consuming Programs Via Batch Jobs	HELP BATJOB
TELENET & Phone Connection Problems	HELP TELENET

To display this menu again:

HELP ME

**@HELP FIND-FILES**

NOTE: Detailed information on sequence file names for the various databases can be found on pages 24-29 in the Introduction to BIONET (BIONET users) or pages 32-38 in the IntelliGenetics Timesharing system manual (commercial users). Please note the following revision to the section on Protein Sequences in those manuals. The file <IG>SNBRFPEP.LIST has been changed to <IG>SNBRF.LST in conformance with the naming conventions for the other analogous database files.

**FOR INFORMATION ON:**  
-----**TYPE AFTER THE @**  
-----

How to use and interpret \*'s in filenames

HELP FF-WILD

Finding a file that contains the sequence  
of a specific gene or protein

HELP FF-FIND

Choosing which files to search in the IFIND  
and QUEST programs using:

the GenBank database (for nucleic acids)  
the EMBL database (for nucleic acids)  
the NBRF database (for proteins)

HELP FF-XNA  
HELP FF-XNA  
HELP FF-PRO

To display the main menu again:

HELP ME

# III. XMULTAN ALIGNMENT

<b>CONSENSUS</b>	<b>PLPAATL TYCANTTTT TTGC JAAALMTGTTTT</b>
<b>1</b>	<b>GTAATA TCCACTTTT TTGC AGAGTCTGTTTT</b>
<b>2</b>	<b>GTAATA TCAACTTTT TGGC AAAATCCGTTTT</b>
<b>3</b>	<b>GAAAT TTCATTTTT TTGC CAAAAGTATTTT</b>
<b>4</b>	<b>AAAAATATTTTAATTT TGGC AAAATCCGTTTT</b>
<b>5</b>	<b>GAAAT TTCATTTTT TTGC CAAAATATTTT</b>
<b>6</b>	<b>GAAATA TCCGATTTT TTACAAAATTTTTTTT</b>
<b>7</b>	<b>GATGATTT TTGGATTTTGT C CAAAATGGATT</b>

**C TYC APATTTTGTCTPTMAAAAAAATAATCAGT TTTTTPYCANAAACAT**

<b>1</b>	<b>TCC AAATTTTCGGTCATC</b>	<b>AAATAATCATT TATTTTGGCACAACAT</b>
<b>2</b>	<b>TCC AAATTTTCGGTCATC</b>	<b>AAATAATCAGTGTCTTCTGCTACACTT</b>
<b>3</b>	<b>CCC AGATTTT TTGTGAAAAAATATTTGGT</b>	<b>ATTTTGTCAAAACAT</b>
<b>4</b>	<b>TTC AAATTTTGATCAT</b>	<b>AAATAATCAGT TTTTGTCAACAACCTC</b>
<b>5</b>	<b>CCC AGATTTT TTGTGAAAAAATATTTGGT</b>	<b>TTTTTGTCAAAACAT</b>
<b>6</b>	<b>TTCGATTTTTTGATC</b>	<b>AAAAATAATCCGT TTTTTTAATAACCA</b>
<b>7</b>	<b>TTC TGTTTTTTTCGATTTTAA</b>	<b>TATCAGT ATTATGATCAGAACAT</b>

**C TJPAATAATTGTCTYPAATATGGAAATGTCATACCTCG TTGAGTTCGTAA**

<b>1</b>	<b>AAAAAATAATTGTCTGAATATGGAAATGTCATATCTCA</b>	<b>CTGAGCTCGTAA</b>
<b>2</b>	<b>TAAAAACAATTGTCTGAATATGGAAATGTCATACCTCG</b>	<b>CTGAGCTCGTAA</b>
<b>3</b>	<b>TCGAATAATTACCCAATATGGAAATGTCATACCTCGTTTGAAGTTGTTT</b>	
<b>4</b>	<b>TAAAAATAATTGTCTGAATTCGGAAATGGTATACCTCG</b>	<b>TTGCGTTCGTAG</b>
<b>5</b>	<b>TCGAATAATTACCCAATATGGAAATGTCATACCTCG</b>	<b>TTGAGTTTGT</b>
<b>6</b>	<b>CAAAAATAGTTGTCCAA GTGGAAATGCCATACCTCG</b>	<b>TTGAATTCGTAA</b>
<b>7</b>	<b>TCGAATATTGGTCCAAATATGGAAATGTCATATCTCG</b>	<b>TTGAATTCGTAA</b>

**C YTLAATTYCCAATCGAACTGTGTTCAA AANTTGGAAATTNLA TTTK**

<b>1</b>	<b>TAAATTTCCAATCGAACTGTGTTCAA</b>	<b>AAA TGGAAATTAA TTTT</b>
<b>2</b>	<b>TTAATTTCCAATCGAACTGTGTTCAA</b>	<b>AAA TGGAAATTAA TTTC</b>
<b>3</b>	<b>CTTAATCCCAATCGAATTGCGTTCAA</b>	<b>GTTTTGG AATTCTA GAGG</b>
<b>4</b>	<b>TTTAATTTCCAATCGAACTGTGTTTAA</b>	<b>AGTTGGAACTCTAT TTTT</b>
<b>5</b>	<b>CTTAATCCCAATCGAATTGCATTGCATTCAAGTTTTGAATTCTA</b>	<b>GAGG</b>
<b>6</b>	<b>CAAAATTCCTATCGATCCATGTAC A</b>	<b>TACTTTGAATTCTA ATTG</b>
<b>7</b>	<b>TTAATTTCCAATCGAACTGTGTTTAC</b>	<b>CAA AAAAAATGAATTTTT</b>



**C TTTKCJATTTTTT6CAAATTTT6AT6AT6NTACCCCTTACAAAAAAT6C6**

**1 TT6CCACATTTT6CAAATTTT6AT6ACCCCTCCTTACAAAAAAT6C6**  
**2 TTTGACATAGT6T6CAAATTTT6AT6AT6 TTACAAAATAT6T6**  
**3 GTTTCAGTTTTT6CAAATATT6AT6A TACCCCTTACAAAAAATTC6**  
**4 TTTGCCATTTT6GCAAATTTTAT6AT6TTACCCCTTACAAAAAAT6C6**  
**5 GTTTCATTTTTTCCAAATTTT6AT6AT6GTACCCCTTACAAAAAATTC6**  
**6 TTTGCCATTTTTCGCAAATTTT6AT6AT6GTACCCCTTATCGAAAAT6C6**  
**7 TTTTATTTTTTTCATTTTTT6AT**

**C AAAATTT6MCCAAAAATTAATTTMCLAAATCJKTMAAAA AGT6ATA**

**1 AAAATT6ATCCAAAAATTAATTTCCCTAAATCCTTCAAAA AGTAATA**  
**2 AAAATTT6CCAAAAAATT6ATTTCTCTAAATCCTTGA AAA AGTAATA**  
**3 AAAATTT6GCCAAAAATTAATTTTACAAATCAGTTTAAA AGT6AAA**  
**4 AAAATCGACCCAAAAACGAATTTCCC AAATCCGTCAAAA AGT6ATA**  
**5 AAAATTT6GCCAAAAATTAATTTTACAAATCAGTTTAAA AGT6AAA**  
**6 AAAATTT6TCAATTTTTTTTTTTT6CGAAAATC GAAAAAGTAGGGATA**

**C GGGATNGTTAGCANT6GTLATTA6CL6CLCAAAACAGTMNTTCTTTYAKC**

**1 GGGATCGTTAGCACT6GTAATTAGCT6CTCAAAACAGATATTCGTACATC**  
**2 GGGATCGTCAGCACT6GTAATTAACT6CTCAAAACAGTTTTTTCATGCATC**  
**3 GGGGTTGTTAGTATT6GTTGTAGGAGTACAAAT6GTACTTCTTTT GC**  
**4 GGGATCGTTAGCATT6GTAATTAGCT6CTCAAAACAGTTATTCTTTTCATG**  
**5 GGGGTTGTTAGTATT6GTTGTAGGAGTACAAAT6GTACTCCTTTTTT6C**  
**6 GACATAGTTAGCTATCTTTATTAGCAGCACAACAGTCTTTATTTTAGC**

**C TNTNTGACCATTTTTAGCCAAAGTTATPPCMAAAA**

**1 TAT6TGACCATTTTTAGCCAAAGTTATACGAAAATTTTC6TTT**  
**2 TATATGACCCTTTTTAGCCAAAGTTATGACAAAAATTTTC6TTT**  
**3 TCTCTGACCATTTTTAGTCAAGTTATAGCCAAAACAGCCAAATT**  
**4 TTTATGATCATTTTTAGCCAAAGTTATGATTAAATGCCAATA**  
**5 TCTCTGACCATTTTTAGTCAAGTTATAGCCAAAAAAGCCAAATT**  
**6 T6TGCGTCCATTTTTAACCAAGTTATG6CCAAAACGCCTATT**

**APPENDIX : Best alignment of seven Drosophila Satellite DNS's as determined by XMULTAN on the BIONET resource. The satellite sequences are from the 1.688 g/cm<sup>3</sup> satellite DNA cloned from 1) D. melanogaster, 359 bp repeat, 2) D. melanogaster, 353 bp repeat, 3) D. mauritiana, 4) D. oreana, 5) D. simulans, 6) D. yakuba, 7) D. teissieri.**

## IV. PROGRAMS CONTAINED ON THE IBM AND MAC DIRECTORIES

**ARC.DOC** Documentation for the ARCHIVE program. This program can pack many files into \*.ARC files and also extract files from these compacted forms. This aides in downloading multiple file packages. This file is also contained in the ARC51.COM file below.

**ARC51.COM** Down load this binary file and then run ARC51. This program unpacks itself resulting in the generation of ARC.EXE and ARC.DOC version 5.1. Then delete ARC51.COM. ARC.EXE can then be used to unpack \*.ARC archive files.

**ARCE.ARC** ARCE is a much smaller version of ARC that can extract files from \*.ARC archives but cannot pack files. It is useful if you only download \*.ARC and never need to prepare \*.ARC archives.

**BINHEX.BAS** BINHEX is a 189 line basic program that can convert binary files to hex files and back again for up and downloading over seven bit data paths. The program HC (hex convert) is much faster and is preferred.

**BINHEX.HLP** Help documentation for BINHEX.BAS.

**EMACS.ARC** Archive file containing a micro computer version of EMACS. Although only a limited subset of EMACS commands are available, this program can edit several files simultaneously and can edit very large files.

**EMACS.DOC** Documentation for microcomputer EMACS. Also contained in EMACS.ARC.

**HC.ARC** Archive file containing HC.COM and HC.DOC. See below.

**HC.COM** Binary file for HC program that converts \*.EXE and \*.COM programs into hexadecimal HEX files or converts HEX files back into \*.EXE or \*.COM files. Needed to transfer binary files over seven bit data paths or networks.

**HC.HEX** A HEX version of HC.COM. You must have BINHEX.BAS or a HEXCONV to convert this file back to HC.COM after downloading.

**KERMIT-V120.EXE** An old compact version of KERMIT that will run on all versions of MS DOS and is relatively bug free. IT is particularly good for use on floppy disks where space is limiting. It is missing many features of MSKERMIT but it an excellent terminal emulator.

**LUE210.COM** A program for extracting files in the compressed \*.LBR

library format. Rename this file to LUE.COM after you have downloaded it.

LUE210.DOC	Documentation for LUE120.COM
LUU208.LBR	The complete LUU.COM program and documentation. The LUU Library Utilities allow you to pack and squeeze files into *.LBR format for up and downloading. This program is useful if you intend to upload *.LBR files or maintain them yourself. If you merely want to download and extract files from libraries, then use LUE210.COM instead.
MODEM7.ASM	Part of the assembler code for the parameter section of the public domain MODEM7 program. Assembler programmers can use this file together with DEBUG to alter the default speed and port settings for MODEM7.
MODEM7.COM	A public domain version of MODEM that allows multiple file transfers to and from most other XMODEM programs. The program is set up to send data out COM1 port at 1200 BAUD but these parameters can be changed by program commands or by altering the program itself with the aid of the MODEM7.ASM file above. The terminal emulator of MODEM7 is very poor and setting TERMINAL ADM3 is recommended.
MODEM7.DOC	Manual for use of the MODEM7 program.
NUSQ110.COM	New version of the Unsqueeze program that can unsqueeze *.?Q? files. This program is needed if any of your downloaded files are in the squeezed format. Rename this program to NUSQ.COM or USQ.COM after downloading.
NUSQ110.DQC	Documentation for NUSQ110.COM. You must convert this *.DQC file to a *.DOC file by running NUSQ110.COM before attempting to read it.
SQ129.COM	Program to squeeze files. Useful to compact files before uploading them to minimize communication time and/ or disk space. The program converts *.COM files to *.CQM files, *.DOC files to *.DQC files, *.exe files to *.EQE files etc for many file types. Any file with a Q in the second position of the extension should be assumed to be in squeezed format and should be unsqueezed before using.
SQ129.DQC	Squeezed form of the documentation for SQ129.COM
STRIP.LBR	Program and documentation for converting Wordstar files to standard ASCII files. It strips out the eighth bits that Wordstar leaves and adds normal carriage returns and line feeds. This is necessary before attempting to upload wordstar files to Bionet or before sharing the

files with other word processors. STRIP can also convert TABS in ASCII files to spaces and also convert files with lots of spaces to TABS.

The following programs are available on the MAC: directory:

BINHEX4.BAS	This Basic program when run will produce a Binhex 4.0 application on your disk. This application will convert hex files (*.HEX) files, compressed hex files (*.HCX) and squeezed hex files (*.HQX) to normal Macintosh files.
BINHEX4.DQC	A squeezed form of the Binhex4 manual. Run Binhex4 on this document to generate a MacWrite file of the manual.
BINHEX4.HQX	A Binhex4 compressed HEX file of the Binhex4 application.
BINHEX4.PAS	A pascal program which when run will generate the Binhex4 application. After running BINHEX4.PAS you may delete it keeping only the resulting application.
EMACS.HQX	A microcomputer version of the mainframe EMACS program that runs on the Macintosh. The file must be downloaded and then converted with Binhex4 to generate the application.
FREETERM18.HQX	A public domain terminal emulator and XMODEM file transfer program. This program is a poor terminal emulator (TERMINAL ADM3) but it does a good job of up and downloading files using the MODEM file transfer program. Convert the file FREETERM18.HQX to the Freeterm application using Binhex4.
MAKE-MAKERS.HQX	This program converts an existing application on the Macintosh into two files. The resulting *.PAS file is an Apple Pascal program which when run will regenerate the desired application. The resulting *.BAS file is a Microsoft Basic program which when run will regenerate the application. This program was used to generate the BINHEX4.BAS and BINHEX4.PAS files listed above. The MAKE-MAKERS.HQX file must be converted with Binhex4.
PACKIT.DQC	A Binhex4 version of the Packit Manual. Convert with Binhex4 to obtain a Macwrite document. The file must be converted with Binhex4 to generate the PackIt manual.
PACKIT.HQX	Packit is a program that packs and unpacks several files into one single file for up and downloading purposes. It is similar to the Library and Archive programs for the IBMPC. PACKIT takes several files including applications and documents and puts them into a single

\*.PIT file. After files are packed by PACKIT then they are usually converted to a \*.HGX file by Binhex4 before transmission. IF one downloads a \*.HGX file and then converts it with Binhex4 and it becomes a \*.PIT application, then PACKIT must be used to separate out the individual files before use. The file must be converted with Binhex4 to generate the PackIt application.

PACKITII.HGX      Version II of the PACKIT program does more sophisticated data compression on the packed files. It is compatible with files packed with version I of PACKIT. The file must be converted with Binhex4 to generate the PackIt II application.

The following programs are available on the <PC-SOFTWARE.ZUKER> directory:

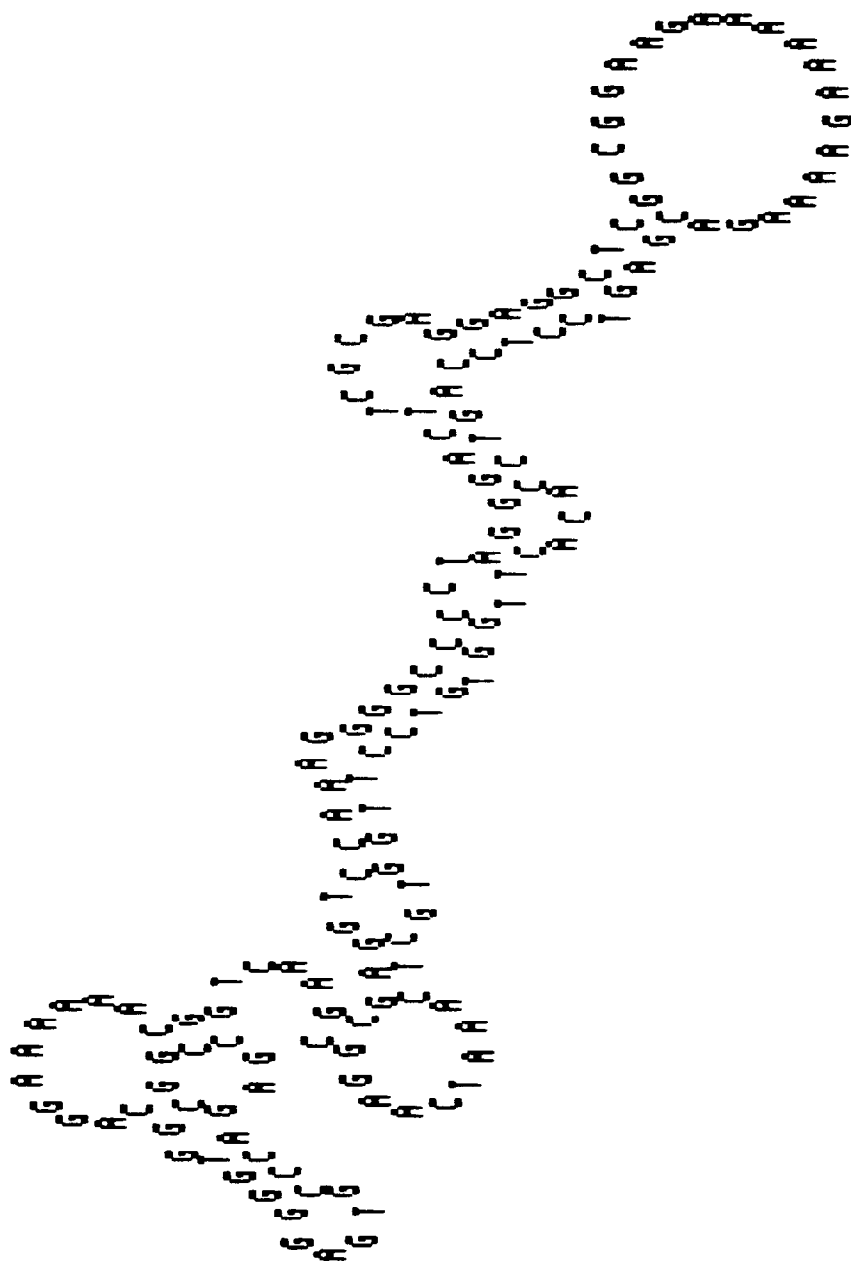
README.DOC		Documentation on this software.
PCFOLD.EXE		Packed version of PCFOLD.
PCFOLD2.EXE		Version without packing of the arrays.
MENUDAT2		Data files needed by the program.
MENUDAT		
MSG	HLP	
FOLD	ENR	Energy file.
FOLD	BAT	Batch file to run the program. This batch file suppresses the 8087 coprocessor before the call to PCFOLD.EXE.
FOLD2	BAT	Batch file to run the program. This batch file suppresses the 8087 coprocessor before the call to PCFOLD2.EXE.
PSTV		Default sequence file.

The directory <PC-SOFTWARE.THOMPSON> contains the following programs:

IMOLECUL.ARC	Programs to display RNA structures graphically using *.CT files produced by Zuker's PCFOLD.
HMOLECUL.ARC	IMOLECUL requires IBM Graphics Card and HMOLECUL requires Hercules graphics card. Files must be extracted with ARC or ARCE.
MOLECULE.DOC	Manual for use of MOLECULE program.

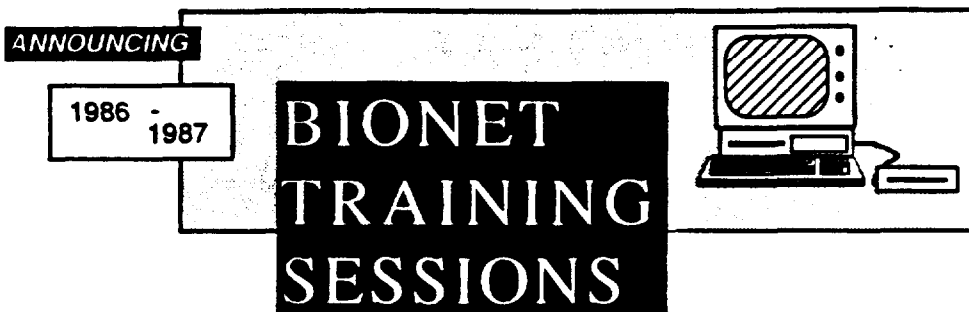
MOLECULE.PAS      Turbo Pascal source code for MOLECULE  
program. MOLECULE requires Turbo Graphix  
Toolkit to be compiled.

## V. DISPLAY BY THOMPSON'S MOLECULE PROGRAM



APPENDIX : RNA SECONDARY STRUCTURE PREDICTED BY PF-FOLD and displayed by J.R. Thompson's molecule program. This figure shows the type of presentation possible with software contributed to the PC-SOFTWARE library. It shows the minimal energy RNA structure of a 150 base pair segment of a viroid RNA predicted by Zuker's PC-FOLD program and displayed by Thompson's MOLECULE program.

## VI. MAILER FOR BIONET TRAINING SESSIONS



IN RESPONSE TO YOUR INTEREST... The BIONET consultants are arranging training sessions in your area. The following training sessions should be of interest to all BIONET users as well as others doing research in the field of Molecular Biology.

**ASBC**

**JUNE 12, 1986 WASHINGTON, D.C. 1:30 TO 4:30 P.M.**

### COURSE CONTENT

#### OVERVIEW OF BIONET AND THE COMPUTER SYSTEM

Core Program Library • Sequence Databases and Development Library • Electronic Mail and Bulletin Boards • System Commands and Directory Organization

#### SEQUENCE DATABASE ORGANIZATION AND SEARCHING METHODS

Rapid Sequence Alignment and Similarity Searches • Sequence Retrieval Using Exact or Ambiguous Patterns

### OTHER CLASSES:

THE FOLLOWING WORKSHOPS HAVE BEEN TENTATIVELY SCHEDULED. CONTENT WILL BE DETERMINED BY YOUR RESPONSE. PLEASE RETURN YOUR REGISTRATION ASAP.

DARTMOUTH COLLEGE, HANOVER, N.H.  
STANFORD UNIVERSITY, PALO ALTO, CA  
UNIVERSITY OF IOWA, IOWA CITY, IA  
MIAMI MID-WINTER SYMPOSIUM

AUGUST 7 - 8  
AUGUST 27, 28, 29  
EARLY SEPTEMBER  
FEBRUARY 11, 1987 (Evening)

Emphasis will be placed on the hand-on use of the Intelligenetics programs as tools for research in Molecular Biology. Depending upon your response, the sessions will be conducted at novice and/or advanced levels. Tuition fees may be necessary to cover costs.

RETURN THIS PORTION TO BIONET CO INTELLIGENETICS 1975 EL CAMINO REAL MOUNTAIN VIEW, CA 94040

Yes, I would like to take advantage of the BIONET workshops. Please send me confirming dates and fee schedule.

Principle Investigator: \_\_\_\_\_  
Sub-Investigator: \_\_\_\_\_  
Institution: \_\_\_\_\_  
Address: \_\_\_\_\_  
City \_\_\_\_\_ ST \_\_\_\_\_ Zip \_\_\_\_\_

I am interested in ☐ NOVICE ☐ ADVANCED  
I am interested in attending the following session

- ☐ DARTMOUTH COLLEGE, HANOVER, N.H.  
☐ STANFORD UNIVERSITY, PALO ALTO, CA  
☐ UNIVERSITY OF IOWA, IOWA CITY, IA  
☐ MIAMI MID-WINTER SYMPOSIUM

☐ I would be interested in hosting a session at my institution

### PLEASE CHECK THE TOPICS THAT ARE OF INTEREST TO YOU:

- |  |   |
|--|---|
| <input type="checkbox"/> Managing large DNA sequencing projects                              | <input type="checkbox"/> DNA or protein sequence analysis |
| <input type="checkbox"/> Restriction mapping tools   | <input type="checkbox"/> Sequence comparison methods      |
| <input type="checkbox"/> Simulation and design of recombinant DNA experiments                | <input type="checkbox"/> Use of BIONET on-line editors    |
| <input type="checkbox"/> DNA or protein sequence database organization and searching methods |   |
| <input type="checkbox"/> Distributed processing: moving data to and from the BIONET computer |   |
| <input type="checkbox"/> Remote job processing - running programs away from the terminal     |   |



## VII. TRAINING SCHEDULES

### **BIONET™ TRAINING SCHEDULE FOR THE MIAMI MID-WINTER SYMPOSIUM**

**BIONET** will be conducting three hands-on training sessions at the Miami Mid-Winter Symposium. **BIONET** subscribers are encouraged to attend the training sessions to become more familiar with the resource and thereby utilize it more fully. Investigators who are not subscribers are invited to learn how **BIONET** can aid in their research. The schedule for the training classes is as follows:

#### **Wednesday, February 5 -- Orange Blossom Room**

##### **1:00 - 2:50 BIONET OVERVIEW**

Programs and features that are available to **BIONET** users, how to access **BIONET**, how to send electronic mail and read or not read bulletin boards, how to manage the files in your **BIONET** account.

##### **ENTERING SEQUENCES**

How to enter and edit nucleic acid and protein sequence data.

##### **3:10 - 5:00 DATABASE SEARCHING**

How to search the NIH and NBRF databases for DNA or protein sequences.

#### **Friday, February 7 -- Orange Blossom Room**

##### **2:00 - 4:00 RESTRICTION ENZYME MAPPING**

How to generate restriction enzyme maps and create and use an individualized restriction enzyme list.

##### **SEQUENCE ASSEMBLY**

How to assemble gel sequence information to generate a consensus sequence.

In order to help cover the expenses of the training a small charge will be made. Cost of the courses is \$30 for one, \$60 for two and \$80 for three. Please enroll at the IntelliGenetics booth #22 or at the Orange Blossom Room.

*The **BIONET** computing resource is a cooperative agreement between IntelliGenetics and the National Institutes of Health. **BIONET** users have access to electronic mail and bulletin boards to connect them to other members of the research community. IntelliGenetics software available on **BIONET** allows users to gather, organize, store, and analyze DNA, RNA and protein sequence data.*

# BIONET™ Training Syllabus

ASBC

Thursday, June 12, 1986

1:30 - 4:30 pm

## BIONET OVERVIEW

Programs and features that are available to BIONET users

- Overview of the core programs
- Databases on line
- Contributed software
  - XFASTP
  - XFASTN
  - IDEAS
  - BIOFLD

- Communications facilities
  - e-mail
  - bboards

## Accessing BIONET

- Using UNINET to connect to BIONET
- Logging in procedures
- Solutions to some common problems associated with log-in
  - UNINET error codes
  - Connected okay, but message is garbled
  - Backspace doesn't work

## Sending electronic mail and using the bulletin boards

- Basics of using the mail program
  - Sending mail /editing messages
  - Reading mail
  - Deleting old messages
- Basics of the BBOARDS
  - Getting a listing of the Bboards
  - Accessing a particular bboard
  - Reading messages
  - Returning to your mail file

## File management

- Viewing files in your account
  - Use of wildcards with the DIR command

- Removing old files
- Archival and retrieval of files
- Important files to keep and files that can be removed
- How to see your disk space and disk usage
- Reallocation of disk space

## **SEQUENCE ENTRY**

Entering and editing nucleic acid or protein sequences

- Use of the GENED program to enter sequences
  - EDIT line editor
  - ESEQ display editor

## **DATABASE SEARCHES**

Locating a sequence or group of sequences in the nucleic acid or protein sequence databases.

- Use of the FIND program for rapid sequence location
  - Files to use with FIND
    - NIH.LST
    - EMBL.LST
    - SNBRF.LST
- Use of the QUEST program for sequence retrieval
  - When to use QUEST
    - Collection of similar sequences to a file
    - Searching for patterns within sequences

Rapid sequence alignment and similarity searches

- Using IFIND to align your sequence with similar sequences
  - Initiating a search with IFIND
  - READING in a .FND file and using it to align sequences